Du corpus vidéo à l'agent expressif: Utilisation des différents niveaux de représentation multimodale et émotionnelle

Jean-Claude Martin, Sarkis Abrilian, Laurence Devillers, Myriam Lamolle, Maurizio Mancini, Catherine I Pelachaud

▶ **To cite this version:**

Jean-Claude Martin, Sarkis Abrilian, Laurence Devillers, Myriam Lamolle, Maurizio Mancini, et al.. Du corpus vidéo à l'agent expressif: Utilisation des différents niveaux de représentation multimodale et émotionnelle. Revue des Sciences et Technologies de l'Information - Série RIA : Revue d'Intelligence Artificielle, 2006, 20 (4-5), pp.477-498. 10.3166/ria.20.477-498 . hal-03580926

HAL Id: hal-03580926
https://hal.science/hal-03580926v1

Submitted on 28 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Du corpus vidéo à l'agent expressif

Utilisation des différents niveaux de représentation multimodale et émotionnelle

J.-C. Martin* — S . Abrilian* — L. Devillers* — M. Lamolle — M. Mancini** — C. Pelachaud****

** LIMSI-CNRS, BP-133 F-91403 Orsay cedex*

{martin, abrilian, devil}@limsi.fr

***LINC, Université Paris 8, 140 rue de la Nouvelle France 91300 Montreuil*

{lamolle, mancini, pelachaud}@iut.univ-paris8.fr

RÉSUMÉ. *Les relations entre émotions et indices multimodaux ont été peu étudiées dans le cadre d'émotions autres que les émotions de base jouées par des acteurs. Dans cet article nous présentons deux expériences permettant d'étudier ces relations par une démarche d'analyse-synthèse. Nous partons d'interviews télévisées montrant des comportements émotionnels naturels. Un protocole et un schéma de codage ont été définis pour les annoter à plusieurs niveaux (contexte, émotion, multimodalité). La première expérience manuelle a permis d'identifier les niveaux de représentation qui interviennent pour faire rejouer ces comportements par un agent expressif. La deuxième expérience fait intervenir une extraction automatique à partir des annotations multimodales. Une telle démarche permet d'apporter des connaissances sur les relations complexes entre émotions et multimodalité.*

ABSTRACT. *Relations between emotions and multimodal behaviors have mostly been studied in the case of acted basic emotions. In this paper, we describe two experiments studying these relations with a copy-synthesis approach. We start from video clips of TV interviews including real-life behaviors. A protocol and a coding scheme have been defined for annotating these clips at several levels (context, emotion, multimodality). The first experiment enabled to manually identify the levels of representation required for replaying the annotated behaviors by an expressive agent. The second experiment involved automatic extraction of information from the multimodal annotations. Such an approach enables to study the complex relations between emotion and multimodal behaviors.*

MOTS-CLÉS : *émotion, multimodal, corpus vidéo, agent expressif.*

KEYWORDS: *emotion, multimodal, video corpus, expressive agent.*

1. Introduction

De nombreuses recherches en Psychologie ont été effectuées sur la communication verbale et non verbale observée lors d'émotions dans les expressions faciales (Ekman, 1999), ou dans les mouvements expressifs du corps (DeMeijer, 1991, DeMeijer, 1989), (Newlove, 1993), (Boone *et al.*, 1998, Boone *et al.*, 1996, Wallbott, 1998). Ces études étaient basées pour la plupart sur les émotions de base (colère, dégoût, peur, joie, tristesse, surprise) jouées par des acteurs. L'annotation de comportements multimodaux communicatifs plus naturels a pourtant été étudiée dans des extraits télévisés mais sans se focaliser sur les émotions (Kipp, 2004) ou sans utilisation d'un outil d'annotation informatisé. Les corpus de données sur les comportements multimodaux lors d'émotions naturelles sont donc peu nombreux en dépit de l'accord général sur la nécessité de collecter des bases de données audiovisuelles qui mettent au premier plan les expressions naturelles des émotions (Douglas-Cowie *et al.*, 2003).

Les Agents Conversationnels Animés (ACAs) utilisent un large éventail de modalités telles que par exemple la parole, les gestes, et les expressions faciales. Ces modalités fournissent à l'interface homme-machine la possibilité d'exprimer vers l'utilisateurs des expressions d'émotions avec potentiellement différentes stratégies de comportements en combinant les différentes modalités. Le comportement émotionnel et l'expressivité de ces agents animés jouent un rôle central pour l'utilisateur, par exemple pour mettre en oeuvre des systèmes de narration interactive. Mais comment définir la dynamique de chaque modalité et leur combinaison lors de comportement émotionnel? A quel niveau de temporalité et d'abstraction faut-il se situer? Comment être certain que le comportement émotionnel de l'agent sera effectivement perçu par l'utilisateur, notamment pour les comportements émotionnels naturels plus complexes que les émotions basiques actées ? L'extériorisation des comportements non verbaux joue en effet un rôle important dans la perception des émotions (Douglas-Cowie *et al.*, 2005).

L'objectif de notre travail est d'étudier les relations entre comportements multimodaux et émotions naturelles. Notre approche consiste à utiliser pour cela la complémentarité entre les corpus multimodaux et les ACAs. Plus concrètement, il s'agira de piloter un ACA à partir de corpus de données réelles naturelles et riches en comportements émotionnels.

Notre approche consiste en deux phases: tout d'abord, annoter la perception de l'émotion à de multiples niveaux dans des vidéos d'interviews télévisées, ensuite rejouer les émotions perçues par un ACA, à partir des annotations.

Dans la première phase, un corpus vidéo d'interviews télévisées a donc été collecté puis annoté manuellement en proposant des dimensions de représentations pertinentes pour annoter la perception de l'émotion. L'annotation de comportements communicatifs dans des environnements sociaux est très complexe en raison du grand nombre de variables intervenant dans le processus communicatif. Il existe de

nombreux schémas d'annotation : des gestes (Kendon, 1993), (McNeill, 1992), (Calbris, 1990), des expressions faciales (Ekman *et al.*, 1978), du regard (Poggi *et al.*, 2000), et des émotions (Cowie *et al.*, 2000) (Ekman, 1999), (Scherer, 2000). Ces schémas sont extrêmement riches en terme de données codées et sont très complexes à utiliser. L'observation de données naturelles nous a permis d'identifier les différents niveaux de codage à retenir pour l'étude des relations entre émotions réelles et comportements multimodaux. Cela nous a également permis de mettre en avant des combinaisons d'émotions : émotions masquées, conflictuelles, etc. La possibilité de définir plusieurs étiquettes émotions pour un même segment émotionnel a ainsi été ajoutée à notre schéma (Abrilian *et al.*, 2005a).

La seconde phase de l'étude consiste à animer un ACA. Plusieurs modèles ont été proposés pour la sélection et l'animation du comportement d'un agent (Cassell *et al.*, 2000a) (Prendinger *et al.*, 2004). Les travaux sur la sélection du comportement concernent principalement les aspects sémantiques des gestes humains (McNeill, 1992). Dans (Cassell *et al.*, 2001), des comportements non verbaux appropriés sont sélectionnés pour accompagner le texte sur la base d'une analyse linguistique. Récemment, ce groupe a étudié la génération de gestes iconiques à partir d'un modèle paramétrique basé sur des analyses de corpus vidéo (Tepper *et al.*, 2004). Noot et Ruttkay soulignent le besoin de variabilité inter sujets dans GESTYLE (Noot *et al.*, 2004), traitant des comportements atomiques basés sur les « dictionnaires de style ». L'animation des gestes nécessite la génération de mouvements réalistes des bras et des mains. Les systèmes d'animation utilisent souvent un langage de représentation personnalisé pour décrire les gestes (Hartmann *et al.*, 2002), (Kopp *et al.*). EMOTE (Chi *et al.*, 2000) implémente un modèle de gestes adaptables pour l'ajout d'informations liées à l'expressivité.

Les résultats provenant de la littérature en Psychologie sont très utiles pour la spécification d'Agents Conversationnels Animés, mais n'apportent à ce jour que peu de détails, et ne traitent pas des variations concernant les facteurs contextuels du comportement multimodal émotionnel. Peu de chercheurs ont utilisé des corpus multimodaux spécifiques au contexte pour la spécification d'un ACA (Kipp, 2004). Dans (Cassell *et al.*, 2000b), les comportements multimodaux de sujets décrivant une maison ont été annotés et utilisés pour spécifier des règles de comportement de l'agent REA.

Dans le domaine des ACAs affectifs, la plupart des travaux effectués à ce jour utilisent des données jouées par des acteurs. La majorité des travaux dans ce domaine de recherche utilise soit la capture de mouvement (Cao *et al.*, 2004), (Egges *et al.*, 2002), (Kshirsagar *et al.*, 2001), soit des vidéos (Tsapatsoulis *et al.*, 2002), (Pandzic, 2003).

Nous différencions notre approche des leurs dans la mesure où nous utilisons un corpus d'émotions naturelles. Notre but est non seulement de reproduire des comportements multimodaux avec un ACA mais surtout d'étudier les relations entre modalités lors de comportements émotionnels, plus particulièrement dans le cas

d'émotions complexes. Nous modélisons ce qui est visible ; c'est-à-dire que nous considérons les signaux perçus et nous essayons de les générer. Les processus qui ont permis d'arriver à la génération de tel ou tel signal ne sont pas pris en compte, seule la partie extériorisée est modélisée. Notre objectif est donc de comprendre de quelle manière une émotion donnée peut être perçue et exprimée, aussi bien quantitativement que qualitativement. Les ACAs nous permettent de synthétiser les indices perçus au niveau multimodal, contexte et émotion.

Nous décrivons dans cet article nos schémas d'annotation, nos protocoles d'extraction d'indices et enfin de génération de comportements multimodaux expressifs. Le schéma de représentation multi-niveaux utilisé à la fois en analyse et en synthèse nous permet d'explorer plusieurs simulations possibles. Deux expériences de simulation ont été ainsi explorées, l'une utilisant une recopie manuelle des indices, l'autre à partir d'indices extraits et calculés sur les annotations du corpus.

Dans ces deux expériences, notre ACA prend en entrée l'annotation effectuée lors de la première phase et calcule les animations du visage et des gestes de l'agent. Le corpus composé de données vidéo très expressives extraites d'interviews de journaux télévisés a été annoté à différents niveaux : émotion, indices multimodaux et contexte, cela avec plusieurs niveaux de granularité et de temporalité. Les trois types d'annotation sont utilisés pour la synthèse avec l'agent, l'émotion pour le comportement expressif global, le contexte et notamment les buts communicatifs (« critiquer », « se plaindre ») pour le modèle de comportement de l'agent, les indices multimodaux et leurs mesures d'expressivité pour la spécification de l'agent.

Cette boucle perception/production offre à terme un moyen de valider notre approche. L'objectif est d'étudier les indices perçus et leurs combinaisons pour la perception d'une émotion donnée. L'utilisation d'un ACA permet d'activer ou de désactiver des signaux donnés. En étudiant ce que les sujets perçoivent à partir de l'animation synthétisée, nous pouvons délimiter les indices les plus pertinents d'une émotion donnée. D'autre part, rejouer des comportements émotionnels avec un ACA permet d'affiner notre schéma d'annotation ainsi que le modèle d'animation, plus précisément le modèle d'expressivité d'un ACA. La complémentarité entre agents expressifs et corpus d'émotions naturelles permet d'approfondir la compréhension des interactions entre les différentes modalités mises en jeu lors de comportements émotionnels.

Dans la section suivante, nous décrivons brièvement un exemple illustrant notre approche. Les sections suivantes détaillent ces différentes étapes. La section 3 décrit les approches utilisées classiquement pour représenter des comportements émotionnels. La section 4 décrit notre phase d'annotation. La section 5 décrit le système d'agent expressif que nous utilisons. Les section 6 et 7 décrivent les deux expériences que nous avons effectuées.

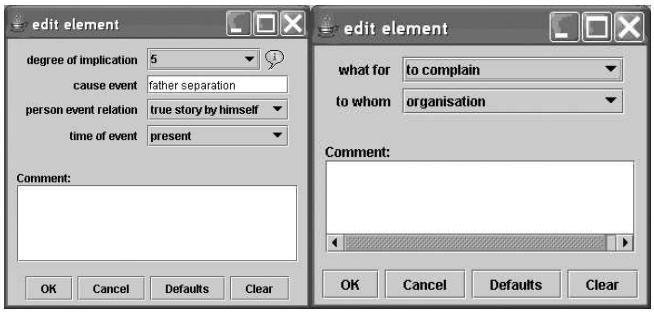
2. Exemple illustratif

L'image gauche de la Figure 1 provient d'un extrait vidéo du corpus EmoTV d'interviews télévisées (Abrilian *et al.*, 2005a). La femme réagit à une épreuve douloureuse récente durant laquelle son père a été emprisonné. Telles que le révèlent les annotations de cet extrait par trois annotateurs, le comportement de cette femme est perçu comme une combinaison complexe de colère et de désespoir, avec la présence de variation temporelle de l'intensité dans l'extrait vidéo. De plus, le comportement émotionnel est perçu à la fois dans la parole et dans plusieurs modalités visuelles (mouvement de tête, regard, mouvement du torse, et gestes de la main). Plusieurs niveaux d'annotation sont effectués dans EmoTV à l'aide de l'outil Anvil (Kipp, 2001) (Figure 2) : certaines informations concernent la vidéo globale (appelé le niveau « global ») ; tandis que d'autres sont liées aux segments émotionnels (le niveau « local ») ; au plus bas niveau, nous retrouvons l'annotation temporelle des comportements multimodaux incluant la qualité du mouvement.

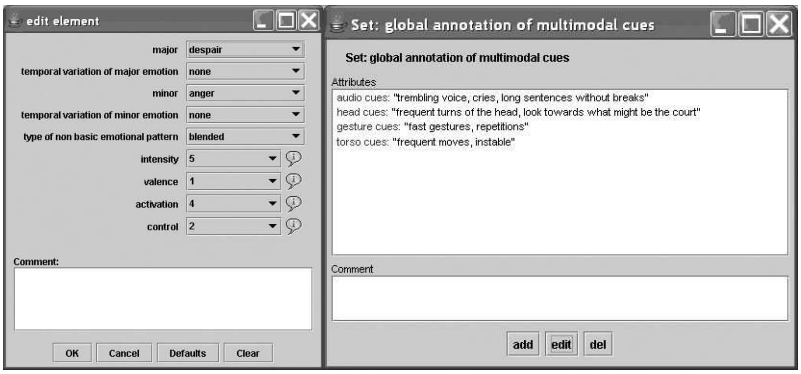
Les étiquettes émotions annotées ainsi que la description de la qualité du mouvement ont été utilisées en entrée du système ACA appelé Greta. Nous avons retranscris la parole de la femme et utilisé cette transcription comme point de départ. La transcription est complétée par des balises permettant de piloter l'animation de l'ACA. L'approche suivie est une boucle analyse-synthèse pour affiner l'animation de l'ACA. L'annotation du segment vidéo est réécrite de manière à correspondre au langage de spécification de l'ACA, APML (De Carolis *et al.*, 2004). Le schéma d'annotation nous permet de préciser deux émotions pour un même segment. Un segment correspond à une unité émotionnelle, cette mesure est perceptive. Trois annotateurs ont validé la segmentation et annoté avec deux étiquettes émotionnelles chacun des segments, une émotion majeure et une mineure si nécessaire [Abrilian, 2005a #1280]. Toutes les annotations sont ensuite combinées pour donner un vecteur d'émotions pondérées (Douglas-Cowie *et al.*, 2005).

Dans l'exemple de la Figure 1, l'émotion correspond à trois segments successifs, le premier annoté uniquement « colère », le deuxième par le vecteur (colère (0.67), désespoir (0.11), déception (0.11), tristesse (0.11)) et enfin le troisième par (désespoir (0.56), colère (0.33), tristesse (0.11)). A l'heure actuelle, à partir du texte APML et du profil comportemental de l'agent, le système calcule l'animation de l'agent.

Figure 1. *Image provenant d'un extrait du corpus, comportement émotionnel naturel : mélange de colère et de désespoir (à gauche). Une première simulation avec l'agent Greta (à droite)*



(a) Annotation au niveau global de l'extrait vidéo par des informations contextuelles (à gauche) et par l'acte communicatif (à droite)



(b) Annotation au niveau global de l'extrait vidéo: étiquettes d'émotions et dimensions (à gauche), zone de texte libre pour les indices multimodaux (à droite)

(c) Annotation au niveau local d'un segment émotionnel non basique par un des annotateurs, annotation d'une combinaison de 2 étiquettes (colère et déception), annotation des dimensions classiques (intensité, valence, activation, contrôle) et des modalités émotionnelles présentes (tête, yeux, torse, etc.).

Figure 2. *Annotation multi-niveaux de comportements émotionnels dans EmoTV. Les annotations multimodales de bas niveau sont décrits à la Figure 3*

3. Représentation de comportements émotionnels

L'annotation et la modélisation de comportements émotionnels nécessite la représentation des multiples niveaux d'abstraction et de temporalité impliqués dans le processus émotionnel : l'émotion elle-même, une représentation du contexte décrivant ici les buts communicatifs et les comportements multimodaux correspondants.

3.1. *Représentations des émotions*

Les études sur les émotions utilisent plusieurs approches : les dimensions d'évaluation (« appraisal »), les dimensions abstraites, et plus communément les catégories verbales. Ces dernières incluent les étiquettes primaires (colère, peur, joie, tristesse, etc.…) (Ekman, 1999)) et les étiquettes secondaires pour les émotions sociales (e .g. amour, soumission). Plutchik (Plutchik, 1994) a également combiné les émotions primaires dans le but de produire des étiquettes pour les émotions intermédiaires. Par exemple, l'amour est une combinaison de joie et d'acceptation, tandis que la soumission est une combinaison d'acceptation et de peur. Le nombre d'étiquettes nécessaires pour annoter des émotions naturelles peut être élevé par rapport aux émotions de base. La plupart des études sur la modélisation des émotions utilise un ensemble réduit d'étiquettes pour être gérable informatiquement (Batliner *et al.*, 2000). Plutôt que d'utiliser ces ensembles réduits d'étiquettes, certains chercheurs définissent les émotions à l'aide de *dimensions* abstraites continues : Activation-Evaluation (Douglas-Cowie *et al.*, 2003), ou Intensité-Evaluation (Craggs *et al.*, 2004). Mais ces dimensions ne permettent pas d'obtenir une représentation précise des émotions. Par exemple, il est impossible de distinguer entre peur et colère. Finalement, le modèle d'évaluation permet de décrire la perception/production de l'émotion. L'avancée majeure de cette théorie est la spécification détaillée des dimensions utilisées pour l'évaluation d'événements antécédents à l'émotion (agrément, nouveauté, etc.…) (Scherer, 2000).

3.2. *Comportement expressif*

Nous définissons un comportement comme une paire (sens, signal). Ces paires peuvent être élaborées à partir d'analyses de corpus vidéo (Poggi *et al.*, 1996). A un sens donné peuvent être associées différents ensembles de signaux. Par exemple, le sens « emphase » (d'un mot) peut apparaître en même temps qu'un haussement de sourcils, qu'une approbation de la tête, ou qu'une combinaison de ces deux signaux. De la même manière, un même signal peut être utilisé pour porter différents sens, par exemple un haussement des sourcils peut être un signe de surprise, d'emphase, ou encore de suggestion. Un troisième élément caractérise un comportement : la manière de l'exécuter, nous appelons ce paramètre l'expressivité du comportement. Le signal est décrit *statiquement* (e.g. une expression faciale à son « apex », la

forme d'un geste à la phase d'exécution). Le paramètre expressif se réfère à la variation dynamique de ce comportement autour de cette description statique, par exemple, la durée temporelle et sa force.

Six dimensions représentant l'expressivité du comportement sont définies dans les études sur les comportements multimodaux émotionnels (Wallbott, 1998), (Gallaher, 1992). Les dimensions d'expressivité ont été conçues pour les comportements communicatifs uniquement. Chaque dimension agit différemment pour chaque modalité. Pour le visage, les dimensions d'expressivité agissent principalement sur l'intensité de la contraction musculaire et sa temporalité (à quelle vitesse un muscle se contracte). Dans le cas du geste, l'expressivité fonctionne au niveau des phases du geste : par exemple la phase de préparation, l'exécution (stroke), la retenue (hold), et au niveau de la coarticulation entre plusieurs gestes. Nous considérons les six dimensions d'expressivité qui ont été identifiées dans la littérature (Hartmann *et al.*, 2005). Trois d'entre elles, expansion spatiale, expansion temporelle, et force agissent sur un geste donné et sur l'expression faciale : respectivement l'amplitude du mouvement (correspondant à un déplacement physique d'un muscle facial ou d'une main), la durée du mouvement (liée à la vitesse d'exécution du mouvement), et les propriétés dynamiques du mouvement (valeur d'accélération du mouvement). Une autre dimension, la fluidité, agit sur plusieurs comportements d'une même modalité et spécifie la fluidité avec laquelle deux comportements s'enchaînent. Les deux dernières dimensions, l'activation globale, et la répétition, jouent sur la quantité et la répétition d'un comportement.

4. Annotation multi-niveaux d'émotions non basiques

Dans une perspective d'étude des comportements multimodaux apparaissant lors de données naturelles (non jouées par des acteurs), les données issues de la littérature peuvent être avantageusement complétées par des corpus audio visuels. Le but du corpus EmoTV est d'apporter des connaissances sur les relations entre modalités lors de comportements naturels émotionnellement riches. Ce corpus comporte 50 extraits vidéo d'interview télévisés contenant des comportements émotionnels (Abrilian *et al.*, 2005a). La difficulté principale dans la définition du schéma de codage permettant d'annoter et de représenter de tels comportements émotionnels réalistes, est de trouver les niveaux de description utiles en terme de granularité et de temporalité. Les spécificités du schéma de codage multi-niveaux utilisé pour EmoTV (que nous détaillons dans cette section) sont : la possibilité d'annoter les étiquettes d'émotion ainsi que les dimensions abstraites dans le but d'étudier leur redondance et complémentarité ; la définition de combinaisons d'émotions non basiques ; l'utilisation de deux étiquettes pour annoter un seul comportement émotionnel ; le contexte émotionnel incluant en autres les but communicatifs et des dimensions basées sur les théories d' « appraisal » ; une description temporelle de la variation d'intensité de l'émotion dans chaque segment ; une description globale des signes d'émotions perçus dans les différentes

modalités ; et une description plus détaillée des comportements multimodaux de chaque segment (Abrilian *et al.*, 2005b).

Dans le but de trouver une liste appropriée d'étiquettes d'émotions, différentes stratégies peuvent être utilisées (Cowie, 2001), (Craggs *et al.*, 2004). Dans EmoTV, deux annotateurs experts ont affecté une étiquette de leur choix (texte libre) à l'émotion perçue dans chaque segment émotionnel. Ainsi, 176 étiquettes ont été utilisées, classifiées ensuite en 14 catégories : colère, désespoir, dégoût, doute, exaltation, peur, irritation, joie, neutre, douleur, tristesse, sérénité, surprise et inquiétude. Nous avons cependant conservé les différents niveaux de granularité de ces catégories. Le haut niveau est composé des 6 classes bien connues d'Ekman (Ekman, 1999) plus les classes « neutre » et « autre ».

Le schéma de codage contient également deux dimensions abstraites classiquement utilisées dans l'étude des émotions (Cowie *et al.*, 2000) : activation (passif, normal, actif) et valence (négatif, neutre, positif). L'intensité (faible, normale, forte) et le contrôle (contrôlé, normal, incontrôlé) apparaissent également dans le schéma de codage car ils fournissent des informations pertinentes pour l'étude des émotions naturelles. De plus, des descripteurs temporels de la variation de l'intensité dans chaque segment sont présents car cette dynamique est très pertinente pour l'animation d'ACA.

En ce qui concerne l'annotation de comportements multimodaux (Figure 3), la transcription de la parole, incluant les marqueurs d'événements non verbaux, a été effectuée en utilisant la norme de transcription LDC, pour Linguistic Data Consortium. Les indices prosodiques et spectraux sont automatiquement extraits. Dans les extraits vidéo, seule la partie supérieure du corps des sujets est visible.

Le schéma de codage contient des pistes pour annoter les poses et les mouvements pour : le torse, la tête, les épaules, les expressions faciales, les gestes des mains. Les annotations entre poses et mouvements alternent. La piste pose de tête contient un attribut pour la position principale (adapté du schéma de codage FACS) : face, tournée vers gauche/droite, inclinée vers gauche/droite, haut/bas, avant/arrière. Une position secondaire est aussi utilisée pour représenter des combinaisons de positions (tête en bas et à droite).

En ce qui concerne l'annotation des gestes, nous avons conservé les attributs classiques (Kipp, 2004) mais en nous focalisant sur les besoins spécifiques pour l'annotation de comportements émotionnels ; par exemple gestes répétitifs et manipulateurs apparaissaient fréquemment dans notre corpus. Notre schéma de codage permet d'annoter les phases structurelles des gestes (McNeill, 1992) : préparation (bras et mains se mettant en position avant exécution du geste), exécution (la partie la plus énergétique), séquence d'exécutions, retenue (juste avant ou après l'exécution), et la rétraction (retour à la position de repos).

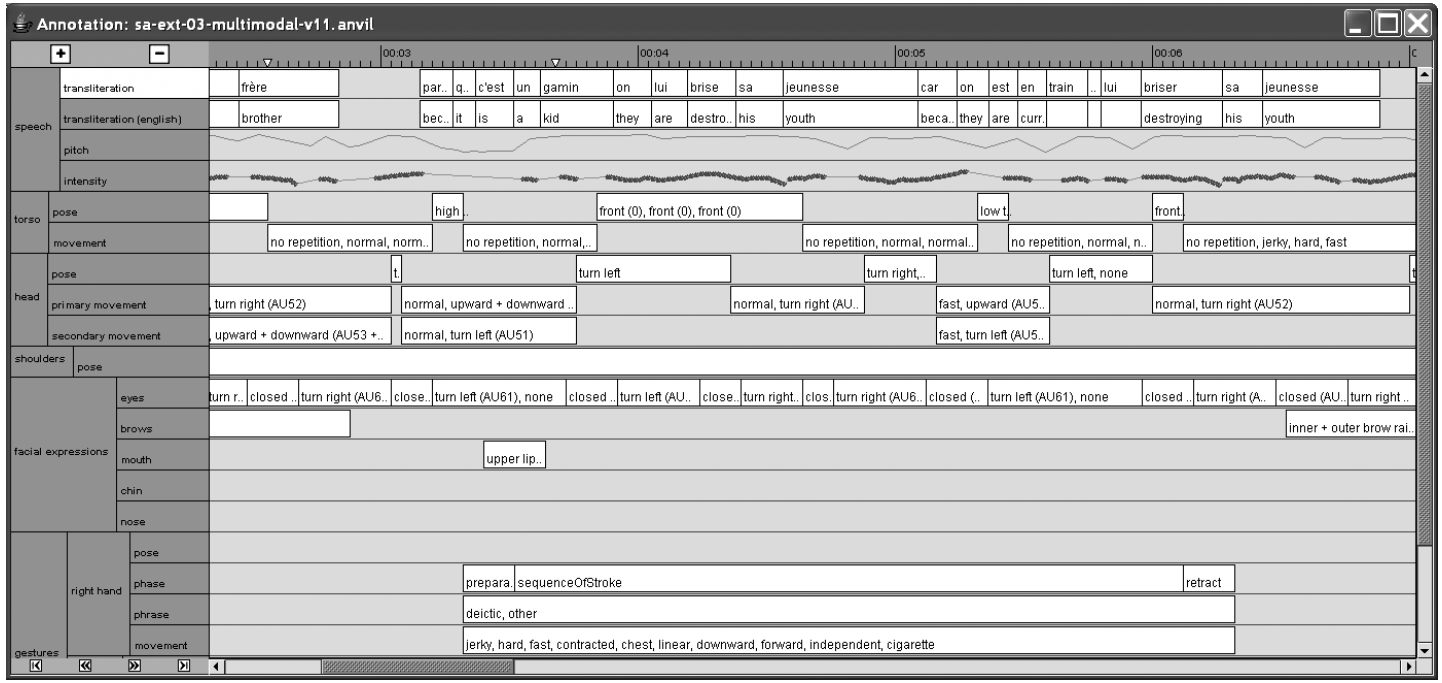


Figure 3. *Annotation des comportements multimodaux émotionnels, de haut en bas : parole (1), alternance de poses et mouvements pour le torse (2) et la tête (3), expressions faciales (les yeux sont souvent fermés dans cet exemple) (4) et gestes de la main incluant les phases structurelles, la fonction et l'expressivité du mouvement*

Nous avons sélectionné les valeurs suivantes pour la fonction du geste : manipulateur (contact avec une partie du corps ou un objet), rythmique (geste synchronisé avec l'emphase dans la parole), déictique (le bras ou la main sont utilisés pour pointer un objet existant ou imaginaire), représentationnel (représente les attributs, actions, relations entre objets et personnages), emblème (mouvement ayant un sens précis et culturellement spécifique). Les gestes représentationnels et les emblèmes sont peu nombreux dans notre corpus, comme l'a montré la première phase d'annotation.

L'expressivité du mouvement, révélée comme très pertinente pour la perception des émotions dans les différentes études citées plus haut, est annotée pour le torse, la tête, les épaules, et les gestes. Les attributs expressifs que nous avons sélectionnés sont : le nombre de répétitions, la fluidité (fluide, normale, saccadé), la force (faible, normale, forte), la vitesse (lente, normale, rapide), l'expansion spatiale (contractée, normale, étendue).

5. Spécification multi niveaux d'émotions et de comportements dans les ACAs

Nous avons développé un système générant les comportements d'un ACA parlant. Notre système considère plusieurs éléments pour le calcul de l'animation finale de l'agent : ce que l'agent vise à communiquer ainsi qu'une description du comportement de base de l'agent.

Notre système d'ACA se base sur une taxonomie de fonctions communicatives proposée par Isabella Poggi (Poggi *et al.*, 2000). Une fonction communicative est définie par un couple (sens, signal) où le sens correspond à la valeur communicative que l'agent souhaite communiquer, et le signal correspond au comportement utilisé pour porter ce sens. Avant elles étaient définies comme un ensemble de buts et de « croyances » que l'utilisateur à l'intention de communiquer. Dans la taxonomie, les fonctions communicatives sont différenciées en informations sur les « croyances » de l'utilisateur, ses intentions, son état affectif, et en information métacognitive sur son état mental. Pour contrôler l'agent, nous utilisons un langage de représentation appelé APML pour Affective Presentation Markup Language, dans lequel les balises sont des fonctions communicatives (De Carolis *et al.*, 2004). Nous avons ajouté une variable *exprFactor* spécifiant l'expressivité avec laquelle un acte communicatif doit être traité. Un exemple de texte (en gras) balisé avec des fonctions communicatives est donné ci-dessous:

```
<performative type="criticize">
<rheme affect="anger" exprFactor="1.3">
Ils ont pris mon père
<boundary type="HL"/>
le jour où ils l'ont pris
<boundary type="LH"/>
ils m'ont pris
<emphasis x-pitch-accent="Lstar" deictic=
```

```
"self3"> moi
…
```

Pour permettre la génération d'un ACA expressif spécifique, nous associons à chaque agent : un profil comportemental spécifiant la hiérarchie des modalités dans lesquelles l'agent est le plus expressif (un agent peut par exemple utiliser principalement ses mains ou son visage pour communiquer (Theune *et al.*, 2004)) ; les prédispositions de l'agent (à quel point chaque modalité est expressive, par exemple un agent peut être plus expressif avec son visage qu'avec ses gestes) ; et l'expressivité globale (caractéristiques globales du comportement de l'agent défini à l'aide de 6 dimensions (Section 4.2)). Le profil comportemental représente en quelque sorte la base (Batliner, communication personnelle) d'un agent ; correspondant à son comportement neutre. Ainsi, les balises APML et les valeurs d'expressivité locale associées modulent cette base. En conséquence, des agents définis avec des profils comportementaux différents réagiront différemment à une même balise APML.

6. Expérience 1 : Identification des niveaux de représentations nécessaires

Afin d'identifier les niveaux de représentation devant intervenir dans la mise en correspondance entre le corpus vidéo et l'agent, nous avons commencé par une première simulation ne faisant pas intervenir de traitement automatique sur les annotations effectuées mais plutôt une définition manuelle des entrées utilisées par l'agent (Figure 4). Ces spécifications sont utilisées par le système pilotant l'agent a deux niveaux : 1) au niveau du texte que doit synthétiser vocalement l'agent et balisé avec le langage APML, et 2) au niveau du profil comportemental de l'agent. Les balises APML, correspondant au sens d'une fonction communicative donnée, sont converties en signaux (faciaux, gestuels). Le système doit tout d'abord sélectionner les comportements les plus appropriés pour un acte communicatif particulier et un agent donné. Il y a deux phases de sélection (Maya et al., 2004) : sélection de la modalité et présélection des signaux. La première sélection détermine la modalité que l'agent utilise ; la seconde consiste à ordonner l'ensemble des comportements possibles ayant un sens similaire. Cet ordonnancement prend en compte l'expressivité de l'agent. Les sorties du système sont les fichiers d'animation et audio pilotant le modèle facial.

7. Expérience 2 : extraction automatique des annotations multimodales

La première expérience décrite dans la section précédente nous a permis d'identifier les niveaux de représentation à considérer pour pouvoir annoter et rejouer les comportements multimodaux émotionnels.

Figure 4. *Etapes utilisées dans la 1ère expérience de simulation : spécification manuelle effectuée en étudiant la vidéo et les annotations des émotions et des gestes dans une vidéo montrant un comportement émotionnel combinant colère et désespoir (à gauche) pour la spécification de l'agent expressif Greta (à droite)*

Les objectifs de cette deuxième expérience étaient d'étudier de manière plus détaillée les possibilités de synthèse plus proches de la vidéo via une utilisation automatique de certaines annotations multimodales de bas niveaux. Ces étapes se regroupent en trois niveaux : annotation, extraction, synthèse (Figure 5).

Nous illustrons notre approche sur un segment émotionnel dont les caractéristiques sont données dans la Table 1. A ce jour, les annotations de la tête et du torse ne sont pas prises en compte pour la génération de l'animation de Greta. Néanmoins elles permettent de compléter le profil d'expressivité : dans le cas du segment 3, 71 segments ont été annotés au total pour les 5 modalités yeux, tête, torse, gestes et sourcils. Ces 71 segments sont répartis en 41 annotations des yeux (57%), 14 annotations de la tête (20%), 7 annotations du torse (10%), 5 annotations des gestes (7%), et 4 annotations des sourcils (6%).

Table 1. Profil expressif d'un segment émotionnel (segment 3 de la vidéo 3)

Extrait vidéo#	**#3**
Durée de l'extrait	37s
Segment	3
Durée of segment	13s
Etiquette d'émotion	Désespoir (56%), colère (33%), tristesse (11%)
Intensité moyenne (1: min – 5 : max)	5
Valence moyenne (1 : négatif, 5 : positif)	1
Nombre de segments multimodaux	71
% mouvement des yeux	57% (41 segments)
% mouvement de la tête	20% (14 segments)
% mouvement du torse	10% (7 segments)
% gestes	7% (5 segments)
% mouvement des sourcils	6% (4 segments)
% rapide vs. lent	15 vs. 0
% fort vs. faible	4 vs. 1
% saccadé vs. fluide	6 vs. 3
% étendu vs. contracté	0 vs. 14

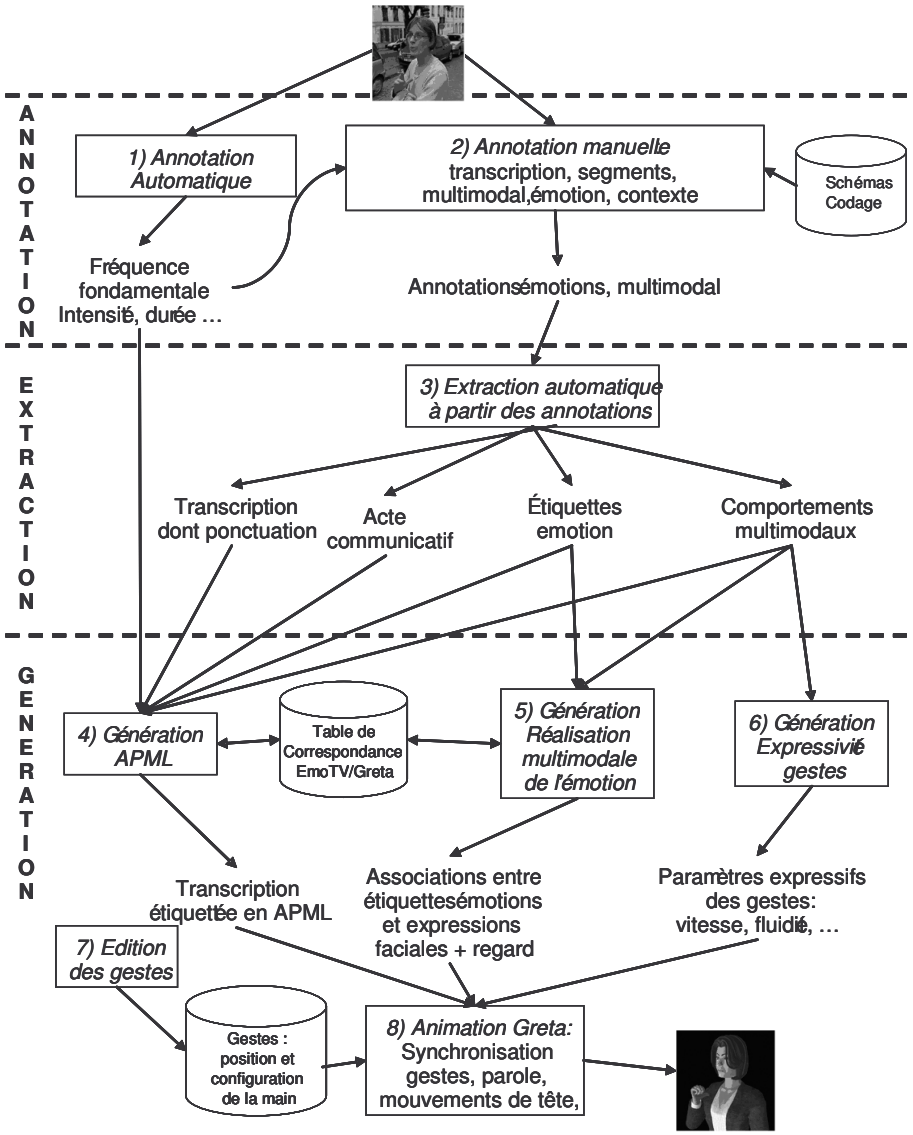


Figure 5. *Etapes utilisées dans la 2ème expérience : détail des niveaux de représentation identifiés pendant la 1ère expérience et extraction automatique de certaines spécifications à partir des annotations*

7.1. *Annotation*

7.1.1. *Etape 1 : annotation automatique*

Cette étape vise à extraire de manière automatique à partir de la vidéo d'origine les informations utiles pour la spécification du comportement de l'agent ou pour l'annotation : fréquence fondamentale, intensité,… Cette étape est réalisée actuellement avec l'outil Praat.

7.1.2. *Etape 2 : annotation manuelle*

Cette étape manuelle effectue plusieurs niveaux d'annotation. La transcription mot à mot avec ponctuation est réalisée en suivant les normes LDC[1] pour les hésitations, les respirations, les rires…

La vidéo est ensuite annotée à plusieurs niveaux temporels (vidéo entière, segments de la vidéo, comportements observés à des instants précis) et à plusieurs niveaux d'abstraction (multimodalité, émotion, contexte). Le comportement global de la séquence est décrit par des annotations sur le contexte, les émotions et les types d'indices multimodaux. Les segments sont annotés avec les émotions, les types de modalités perçues et intègrent aussi une variation temporelle de l'intensité dans le segment. Cette étape fait intervenir les outils Praat et Anvil et les schémas de codage décrits dans la section 5.

7.2. *Etape 3 : extraction automatique à partir des annotations*

Un programme a été développé pour extraire, à partir des différentes annotations effectuées sur la vidéo, les informations qui ont été identifiées durant la première expérience comme nécessaires à la génération : la transcription, l'acte communicatif, les étiquettes et dimensions de l'émotion, les comportements multimodaux (et notamment le nombre d'occurrences et la durée de chaque comportement multimodal pendant chaque segment émotionnel).

7.3. *Génération*

7.3.1. *Etape 4 : génération du fichier APML*

Cette étape consiste à générer le fichier APML utilisé par l'agent à partir des données extraites des annotations, et plus particulièrement des informations audio (transcription et fréquence fondamentale) et de l'acte communicatif. La transcription est directement utilisée dans le fichier APML, car elle correspond au texte que l'agent Greta devra synthétiser vocalement. Elle est complétée par différentes balises. La fréquence fondamentale permet de valider/corriger le passage de la ponctuation de la transcription à l'annotation de contours prosodiques adaptés du modèle ToBi utilisés par APML. Par exemple, la « terminaison » LL correspond à un point de fin de phrase, et LH à une virgule. Nous avons défini une table de

[1] http://www.ldc.upenn.edu/

correspondance associant l'acte communicatif utilisé pour l'annotation et le performatif le plus proche dans l'agent. Ainsi, le but communicatif « se plaindre » utilisé pour annoter l'extrait vidéo 3 du corpus est traduit en performatif « critiquer », correspondant à une spécification au niveau global de l'agent (critiquer = regarder_interlocuteur + froncer_sourcils + bouche_critiquer). Cette performative sera combinée aux valeurs d'expressivité des gestes et des expressions faciales de l'agent, pour obtenir l'animation finale.

7.3.2. *Etape 5 : spécification de la réalisation multimodale de l'émotion*

Dans les vidéos étudiées, les comportements émotionnels sont complexes et sont généralement annotés avec plusieurs étiquettes différentes. Ces annotations effectuées par plusieurs annotateurs sont regroupées en un vecteur d'émotion. Ainsi, le segment 3 de la vidéo 3 a été annotée par le vecteur suivant : 56% de désespoir, 33% de colère et 11% de tristesse. Les deux catégories les plus représentatives de ce vecteur (désespoir et colère) sont regroupées en une étiquette combinée « DespairAnger ». Les annotations des comportements multimodaux observés au niveau du visage durant ce segment sont ensuite utilisées pour associer à cette étiquette une combinaison de comportements des yeux et des sourcils.

Pour les mouvements des yeux, les annotations apportent des informations sur la direction du regard (droite, gauche, haut, bas). Les durées des regards extraits des annotations permettent de pondérer les regards à droite et à gauche, et de spécifier les durées maximums pendant lesquelles l'agent va regarder son interlocuteur ou regarder ailleurs. Dans le cas de notre segment 3, qui a une durée totale de 13 secondes, 41 segments ont été annotés pour les yeux, regard gauche (3 segments, 1.6 secondes (12% du temps)), regard droite (15sgts, 5.84s (45%)). Les résultats montrent que le sujet à également regardé en bas (2sgts, 0.32s (2%)), son interlocuteur (1sgt, 0.03s (0.2%)), et a cligné ou fermé les yeux (20sgts, 3.79s (29%)). Dans notre première expérience de génération d'animation à partir d'annotations, nous avions traité uniquement les annotations droite et gauche, et considéré que le reste du temps (ici 43% du temps), l'agent regarde son interlocuteur. L'utilisation automatique des annotations de bas niveaux permet ici un niveau plus fin et plus fidèle à la vidéo d'origine.

Pour les sourcils, deux informations sont apportées par les annotations, leur mouvement (froncés ou relevés, avec la durée), et leur intensité. Dans le cas du segment 3, 4 segments ont été annotés pour les sourcils, tous avec la valeur froncés (4sgts, 3.4s (26% de 13 secondes)), et avec une intensité forte. Dans l'agent, ces annotations sont traduites en « high_frown », correspondant à un froncement de sourcil de forte intensité pendant 26% (environ 1/4) du segment émotionnel.

Au final, l'affect « DespairAnger» est défini de la manière suivante, sachant que les valeurs avant chaque parenthèse fermante correspondent aux probabilités d'exécution de chacun des comportements :

```
DespairAnger:=

// 12% regard gauche dont 1/4 de sourcils froncés
(eyes_left + high_frown, 0.03), (eyes_left, 0.09),

// 45% regard à droite dont 1/4 de sourcils froncés
(eyes_right + high_frown, 0.11), (eyes_right, 0.34),

// 43% regard vers l'interlocuteur
// dont 1/4 de sourcils froncés
 (look_at + high_frown, 0.10), (look_at, 0.33).
```

7.3.3. *Etape 6 : génération des paramètres expressifs des gestes*

En ce qui concerne les paramètres d'expressivité des gestes, la correspondance est effectuée de la manière suivante : 5 segments on été annotés, d'une durée totale de 11.68 secondes ; les attributs en rapport à la qualité du mouvement, c'est-à-dire la fluidité : fluide (3 segments, 9.16 secondes), la force : forte (2sgts, 2.5s), faible (1sgt, 1.1s), la vitesse : rapide (5sgts, 11.68s), et l'expansion spatiale : contractée (5sgts, 11.68s), déterminent les valeurs des paramètres expressifs correspondant de l'agent Greta. Les durées d'annotations de chaque valeur présente dans la fenêtre temporelle du segment émotionnel sont utilisées pour calculer les valeurs de chacun des paramètres d'expressivité.

Par exemple, dans le cas du 3ème segment, la durée totale du segment étant de 13 secondes, les valeurs obtenues sont les suivantes : pour le paramètre « fluidité », 9.2 secondes ont été annotées avec la valeur « fluide », 2.4 secondes avec la valeur saccadée, et 1.4 secondes avec la valeur « normale »; sachant que le paramètre fluidité (FLT) de Greta est compris entre -1 (saccadé) et +1 (fluide), la valeur affectée est égale à 0.52 (9.2s = 70% fluide, 2.4s = 18% saccadé, 1.4s= 12% normal => 0.70 - 0.18 = 0.52).

7.3.4. *Etape 7 : création des gestes*

Si un geste annoté dans une vidéo n'est pas répertorié dans les gestes gérés par l'agent Greta, il est nécessaire de l'ajouter à l'aide de l'éditeur de configurations de gestes qui permet de spécifier la position et la configuration de la main. Ainsi plus le corpus traité s'agrandit, et plus la bibliothèque gestuelle de Greta s'enrichit de nouveaux gestes.

7.3.4. *Etape 8 : animation Greta*

L'agent Greta prend en entrée plusieurs fichiers : la transcription étiquetée avec APML, les associations entre labels émotions et les expressions faciales, les paramètres expressifs des gestes. Le module gère la synchronisation entre parole, gestes et mouvements de tête. Le résultat est une animation.

8. Conclusions et perspectives

Dans cet article nous avons présenté une approche par analyse/synthèse pour étudier les différents niveaux intervenant dans la réalisation multimodale des comportements émotionnels. Globalement deux niveaux d'indices sont extraits des annotations ; l'un pour simuler le comportement de l'utilisateur se servant des annotations sur les comportements émotionnels et des buts communicatifs, l'autre pour synthétiser ce comportement à l'aide d'indices multimodaux. Les annotations contextuelles contiennent d'autres informations dont des dimensions d' « appraisal » comme l'implication, la nouveauté, etc. qui seraient à terme intéressantes de considérer dans le modèle de comportement de l'agent. De même, la prise en compte d'autres niveaux d'annotations multimodales annotés comme pertinents dans la vidéo (par exemple mouvements de tête) pourraient aussi être considérés dans la génération du comportement de l'agent afin d'envisager différents niveaux de fidélité du comportement de l'agent par rapport à la vidéo d'origine.

Notre approche par analyse/synthèse est particulièrement intéressante pour la création d'agents expressifs basés sur des émotions plus complexes que les émotions de base. Nous envisagerons de valider cette procédure par des tests perceptifs, dans le but d'évaluer à quel point les indices contextuels, l'émotion et les comportements multimodaux sont perceptivement équivalents dans la vidéo originale et dans la simulation du comportement correspondant avec l'agent.

Remerciements

Les travaux présentés dans cet article ont été en partie financés par le Réseau d'Excellence Humaine (Human-Machine Interaction Network on Emotion) IST-2002-2.3.1.6 / Contract no. 507422 (http://emotion-research.net/). Les auteurs remercient également Bjoern Hartmann pour l'implémentation du module générant le comportement expressif et Vincent Maya pour son aide.

Références

Abrilian S., Devillers L., Buisine S., Martin J.-C. "EmoTV1: Annotation of Real-life Emotions for the Specification of Multimodal Affective Interfaces." *Proceedings of the 11th International Conference on Human-Computer Interaction (HCII'2005)*, Las Vegas, Nevada, USA, 22 - 27 July.

Abrilian S., Martin J.-C., Devillers L. "A Corpus-Based Approach for the Modeling of Multimodal Emotional Behaviors for the Specification of Embodied Agents." *Proceedings of the 11th International Conference on Human-Computer Interaction (HCII'2005)*, Las Vegas, Nevada, USA, 22 - 27 July.

Batliner A., Fisher K., Huber R., Spilker J., Noth E. "Desperately seeking emotions or: Actors, wizards, and human beings." *Proceedings of the ISCA Workshop on Speech and*

Emotion: A Conceptual Framework for Research, Newcastle, Northern Ireland, September, p. 195-200.

Boone R. T., Cunningham J. G., "Children's decoding of emotion in expressive body movement: The development of cue attunement." *Developmental Psychology*, vol. 34, n° 5, 1998, p. 1007-1016.

———. "Children's Understanding of Emotional Meaning in Expressive Body Movement." *Proceedings of the Biennial Meeting of the Society for Research in Child Development*, Washington, DC.

Calbris G. *The semiotics of French gestures*. Bloomington Indiana: University Press, 1990.

Cao Y., Faloutsos P., Kohler E., Pighin F. "Real-time Speech Motion Synthesis from Recorded Motions." *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, Grenoble, France, Septembre.

Cassell J., Bickmore T., Campbell L., Vilhjálmsson H., Yan H. "Human Conversation as a System Framework: Designing Embodied Conversational Agents." In *Embodied Conversational Agents*, edited by J.; Sullivan Cassell, J.; Prevost, S.; Churchill, E., pp. 29-63: Cambridge, MA: MIT Press, 2000a.

Cassell J., Stone M., Hao Y. "Coordination and context-dependence in the generation of embodied conversation." *Proceedings of the First International Natural Language Generation Conference (INLG'2000)*, Mitzpe Ramon, Israel, 12-16 June, p. 171-178.

Cassell J., Vilhjàlmsson H., Bickmore T. "BEAT: the Behavior Expression Animation Toolkit." *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH '01)*, Los Angeles, CA, August 12-17, p. 477-486.

Chi D., Costa M., Zhao L., Badler N. "The EMOTE model for effort and shape." *Proceedings of the 27th annual Conference on Computer Graphics and Interactive Techniques (CA'2000)*, Philadelphia, PA USA, 3-5 May, p. 173-182.

Cowie R., "Emotion recognition in human-computer interaction", *IEEE Signal processing Magazine*, vol., n° 18, 2001.

Cowie R., Douglas-Cowie E., Savvidou S., McMahon E., Sawey M., Schröder M. "'FEELTRACE': An Instrument for Recording Perceived Emotion in Real Time." In *ISCA Workshop on Speech & Emotion*, 19-24. Northern Ireland, 2000.

Craggs R., Wood M. M. "A categorical annotation scheme for emotion in the linguistic content." *Proceedings of the Affective Dialogue Systems (ADS'2004)*, Kloster Irsee, Germany, June 14-16.

De Carolis B., Pelachaud C., Poggi I., Steedman M. "APML, a Markup Language for Believable Behavior Generation." In *Life-like characters. Tools, affective functions and applications.*, edited by H. Prendinger and M. Ishizuka, 65-85: Springer, 2004.

DeMeijer M., "The attribution of agression and grief to body movements : the effect of sex-stereotypes", *European Journal of Social Psychology*, vol. 21, 1991, p. 249-259.

———, "The contribution of general features of body movement to the attribution of emotions", *Journal of Nonverbal Behavior*, vol., n° 13, 1989, p. 247 - 268.

Douglas-Cowie E., Campbell N., Cowie R., Roach P., "Emotional speech: towards a new generation of databases", *Speech communication*, vol. 40, 2003.

Douglas-Cowie E., Devillers L., Martin J.-C., Cowie R., Savvidou S., Abrilian S., Cox C. "Multimodal Databases of Everyday Emotion: Facing up to Complexity." *Proceedings of the 9th European Conference on Speech Communication and Technology (Interspeech'2005)*, Lisbon, Portugal, September 4-8, p. 813-816.

Egges A., Kshirsagar S., Magnenat-Thalmann N. "Imparting Individuality to Virtual Humans." *Proceedings of the First International Workshop on Virtual Reality Rehabilitation*, Lausanne, Switzerland, November.

Ekman P. "Basic emotions." In *Handbook of Cognition & Emotion*, edited by T. Dalgleish and M. J. Power, 301–320. New York: John Wiley, 1999.

Ekman P., Friesen W. V. *Manual for the facial action coding system*. Palo Alto, CA: Consulting Psychology Press, 1978.

Gallaher P., "Individual differences in nonverbal behavior: Dimensions of style", *Journal of Personality and Social Psychology*, vol., n° 63, 1992, p. 133–145.

Hartmann B., Mancini M., Pelachaud C. "Formational Parameters and Adaptive Prototype Instantiation for MPEG-4 Compliant Gesture Synthesis." *Proceedings of the Computer Animation (CA'2002)*, Geneva, Switzerland, 19-21 June.

———. "Implementing Expressive Gesture Synthesis for Embodied Conversational Agents." *Proceedings of the Gesture Workshop (GW'2005)*, Vannes, France, May.

Kendon A. "Human gesture." In *Tools, Language and Intelligence*, edited by T. Ingold and K. Gibson: Cambridge University Press, 1993.

Kipp M. "Anvil - A Generic Annotation Tool for Multimodal Dialogue." *Proceedings of the Eurospeech'2001*, Aalborg, Danemark, September.

———. *Gesture Generation by Imitation. From Human Behavior to Computer Character Animation.* Florida: Boca Raton, Dissertation.com, 2004.

Kopp S., Wachsmuth I. "A knowledge-based approach for lifelike gesture animation." *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI)*, Berlin, Germany, August 20-25.

Kshirsagar S., Molet T., Magnenat-Thalmann N. "Principal components of expressive speech animation." *Proceedings of the Computer Graphics International*, February, 2001, p. 38-44.

McNeill D. *Hand and mind - what gestures reveal about thoughts*: University of Chicago Press, IL, 1992.

Newlove J. *Laban for actors and dancers*. New York: Routledge, 1993.

Noot H., Ruttkay Z. "Gesture in Style." In *Gesture-Based Communication in Human-Computer Interaction - GW 2003*, edited by A. Camurri and G. Volpe: Springer, 2004.

Pandzic I. S., "Facial Motion Cloning", *Elsevier Graphical Models Journal*, vol. 65, n° 6, 2003, p. 385-404.

Plutchik R. *The Psychology and Biology of Emotion*. New York: Harper Collins College, 1994.

Poggi I., Caldognetto E. "A score for the analysis of gesture in multimodal communication." *Proceedings of the Workshop on the Integration of Gesture in Language and Speech*, Newark, 7-8 ottobre, p. 235-244.

Poggi I., Pelachaud C., deRosis F., "Eye communication in a conversational 3D synthetic agent", *AI Communications. Special Issue on Behavior Planning for Life-Like Characters and Avatars.*, vol. 13, n° 3, 2000, p. 169-181.

Prendinger H., Ishizuka M. *Life-like characters. Tools, affective functions and applications.*: Springer, 2004.

Scherer K. R. "Emotion." In *Introduction to Social Psychology: A European perspective*, edited by M. Hewstone & W. Stroebe, 151-191: Oxford: Blackwell, 2000.

Tepper P., Kopp S., Cassell J. "Content in Context: Generating Language and Iconic Gesture without a Gestionary." *Proceedings of the Workshop on Balanced Perception and Action in ECAs at Automous Agents and Multiagent Systems (AAMAS)*, New York, NY.

Theune M., Heylen D., Nijholt A. "Generating Embodied Information Presentations." In *Multimodal Intelligent Information Presentation*, edited by O. Stock and M. Zancanaro, 47-70: Kluwer Academic Publishers, 2004.

Tsapatsoulis N., Raouzaiou A., Kollias S., Cowie R., Douglas-Cowie E. "Emotion Recognition and Synthesis based on MPEG-4 FAPs." In *MPEG-4 Facial Animation*, edited by I.S. Pandzic and R. Forchheimer. UK: John Wiley & Sons, 2002.

Wallbott H. G., "Bodily expression of emotion", *European Journal of Social Psychology*, vol. 28, 1998, p. 879-896.